

COMPRESSION OF LOGS OF LANGUAGE DATA

ABSTRACT OF THE DISCLOSURE

A method and apparatus for compressing
5 query logs is provided. Multiple levels of user-
specifiable compression include character-based
compression, token-based compression, and
subsumption. An efficient method for performing
subsumption is also provided. The compressed query
10 logs are then used to train a statistical process
such as a help function for a computer operating
system.